Anonymisierung: papierdünner Datenschutz

Maximilian Steinbeis 2009-09-02T10:57:47

Anonymisieru

schützt Ihre Daten? Denken Sie vielleicht. Denkt auch der Gesetzgeber. Tut sie aber nicht.

Neue Data-Mining-Technologien pusten diese Art Datenschutz weg wie nix. Darauf macht der IT-Rechtler Paul Ohm von der University of Colorado Law School in einem sehr, sehr lesenswerten neuen Aufsatz aufmerksam.

> These advances should trigger a sea change in the law, because nearly every information privacy law or regulation grants a get-out-of-jail-ree card to those who anonymize their data.In the United States, federal privacy statutes carve out exceptions for those who anonymize. In the European Union, the famously privacy-protective Data Protection Directive extends a similar safe harbor through the way it defines “personal data.” Yet reidentification science exposes the underlying promise made by these laws—that anonymization protects privacy—as an empty one, as broken as the technologists’ promises.

Anonymisierte Daten sind überall. Banken, Krankenhäuser, Versicherungen erheben sie und werten sie aus, verkaufen sie teilweise auch weiter. Ganze Wissenschaftsdisziplinen leben von nichts anderem als der Auswertung anonymisierter Daten. Auch ich hier, der kleine Max Steinbeis, erhebe Daten, wer meine Website besucht. Ich kann sehen, über welchen Link Sie kommen, welchen Browser Sie verwenden, wann Sie welchen Blogeintrag aufgerufen haben. Nicht Sie persönlich natürlich. Nicht mit Ihrem Namen oder Ihrer IP-Adresse. Anonymisiert halt.Paul Ohm jedenfalls warnt mit drastischen Prognosen:

> Accretive reidentification makes all of our secrets fundamentally easier to discover and reveal. Our enemies will find it easier to connect us to facts that they can use to blackmail, harass, defame, frame, or discriminate against us. Powerful reidentification will draw every one of us closer to what I call our personal “databases of ruin.”

Was die Möglichkeiten betrifft, den Datenschutz gesetzgeberisch anzupassen und zu verbessern, ist Ohm pessimistisch:

> regulators can protect privacy in the face of easy reidentification only at great cost. Because the utility and privacy of data are intrinsically connected, no regulation can increase data privacy without also decreasing data utility. No useful database can ever be perfectly anonymous, and as the utility of data increases, the privacy decreases.

Was heißt überhaupt Anonymisierung? Dass alle Daten aus der Datenbank verschwinden, mit denen man Rückschlüsse auf die individuellen Personen generieren kann. Wenn man sich eine Datenbank mit "Name", "Geburtsdatum", "Postleitzahl" und "Geschlecht" vorstellt, dann reicht es nicht, die Kategorie "Name" zu entfernen – mit den drei anderen Kategorien kriegt man ebenfalls eine Menge Leute identifiziert. Also muss man die auch entfernen. Dann bleibt von der Datenbank nichts übrig. Die Erkenntnis: Datenbanken sind immer nur entweder nützlich oder anonym, aber nie beides zusammen.

Haut diese Erkenntnis dem gesamten Datenschutzrecht die Füße weg? In den USA ja. Aber in Europa sind die Folgen andere: Das Datenschutzrecht beruht auf der Unterscheidung zwischen personenbezogenen und nicht personenbezogenen Daten. Personenbezogene Daten sind nach § 3 I BDSG

> Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natu#rlichen Person (Betroffener)

und nach Art. 2a EG-Datenschutzrichtlinie. Danach sind personenbezogene Daten

> alle Informationen über eine bestimmte oder bestimmbare natürliche Person; als bestimmbar wird eine Person angesehen, die direkt oder indirekt identifiziert werden kann, insbesondere durch Zuordnung zu einer Kennummer oder zu einem oder mehreren spezifischen Elementen, die Ausdruck ihrer physischen, physiologischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität sind.

Die Rechtsfolge ist also umgekehrt: Durch die neuen Technologien werden vorher nicht bestimmbare Personen zu bestimmbaren Personen – also erweitert sich der Anwendungsbereich des Datenschutzrechts; was zuvor als nicht personenbezogen galt, ist eben jetzt personenbezogen.

Aber diese Definition beruht offensichtlich auf der Annahme, dass man den Daten ansehen kann, ob sie etwas über die "physische, physiologische etc. Identität" einer Person aussagen oder nicht. Genau das wäre dann nicht mehr der Fall: Gibt es gar keine "nicht bestimmbare" Person mehr, dann ist die Unterscheidung, die der Legaldefinition zugrundeliegt, hinfällig.

Ohm beschreibt das so:

> As reidentification science advances, it expands the EU Directive like an ideal gas to fit the shape of its container. A law that was meant to have limits is rendered limitless. A careful balance struck by legislators between privacy and information flow shifts wildly to impose data handling requirements to all data in all situations.

Was tun? Ohm empfiehlt, das "Whack-a-mole-Game" einzustellen ("even if you manage to whack one mole, another will pop right up") und sich von der Unterscheidung personenbezogener und nicht personenbezogener Daten zu verabschieden.Stattdessen sollte der Gesetzgeber sektorielle Risikoanalysen

– wie groß ist die Gefahr, dass tatsächlich jemand die Daten entanonymisiert? Wie schlimm wäre das für die betroffenen Personen? usw. – vornehmen und die Regulierung daran orientieren.

Der EU rät er, auch wenn es auf den ersten Blick absurd klingt, das generelle Datenschutzniveau sogar zu senken:

> the European Union might want to reconsider whether it should lower the floor of its comprehensive data handling obligations. Even if it does not do this (but especially if it does) the EU hould also begin to tackle privacy regulations sectorally much more often than they do today. What might be needed above the comprehensive floor for health records may not be needed for phone records, and what might solve the problems of private data release probably will not work for public releases.

Hat Tip an Concurring Opinions.
